## Contents

*List of figures*  
*List of tables*  
*Preface*

### Part I  Basic topics

1. **In the beginning**  

2. **Basic notions of statistics**  
   2.1 Introduction  
   2.2 Data  
     - 2.2.1 The importance of understanding data  
   2.3 A note on mathematical notation  
     - 2.3.1 Summation  
     - 2.3.2 Product  
   2.4 Probability  
     - 2.4.1 Relative frequencies  
     - 2.4.2 Defining random variables  
     - 2.4.3 Probability distribution functions  
     - 2.4.4 Cumulative distribution functions  
     - 2.4.5 Multivariate probability density functions  
     - 2.4.6 The multivariate probability function  
     - 2.4.7 Marginal probability density functions  
     - 2.4.8 Conditional probability density functions  
     - 2.4.9 Defining statistical independence  
   2.5 Properties of random variables  
     - 2.5.1 Expected value  
       - 2.5.1.1 Properties of expected values  
     - 2.5.2 Variance  
       - 2.5.2.1 Properties of variance

---

*List of figures*  
*List of tables*  
*Preface*
6.7.2 Stratified random sampling 190
6.7.3 Conclusion to the theory of calculating sample sizes 192
6.8 Sampling for choice data: the reality 193

7  NLOGIT for applied choice analysis: a primer 197
   7.1 Introduction 197
   7.2 About the software 197
      7.2.1 About NLOGIT 197
      7.2.2 About NLOGIT/ACA 198
      7.2.3 Installing NLOGIT/ACA 198
   7.3 Starting NLOGIT/ACA and exiting after a session 198
      7.3.1 Starting the program 198
      7.3.2 Inputting the data 198
      7.3.3 Reading data 200
      7.3.4 The project file 200
      7.3.5 Leaving your session 201
   7.4 Using NLOGIT 201
   7.5 How to get NLOGIT to do what you want 202
      7.5.1 Using the Text Editor 202
      7.5.2 Command format 204
      7.5.3 Commands 205
      7.5.4 Using the Project File Box 206
   7.6 Useful hints and tips 206
      7.6.1 Limitations in NLOGIT (and NLOGIT/ACA) 207
   7.7 NLOGIT software 207
      7.7.1 Support 208
      7.7.2 The program installed on your computer 208
      7.7.3 Using NLOGIT/ACA in the remainder of the book 208
   Appendix 7A Diagnostic and error messages 208

8  Handling choice data 218
   8.1 Introduction 218
   8.2 The basic data setup 219
      8.2.1 Entering multiple data sets: stacking and melding 222
      8.2.2 Handling data on the non-chosen alternative in RP data 222
      8.2.3 Combining sources of data 224
      8.2.4 Weighting on an exogenous variable 226
      8.2.5 Handling rejection: the "no option" 227
   8.3 Entering data into NLOGIT 230
      8.3.1 Entering data directly into NLOGIT 230
      8.3.2 Importing data into NLOGIT 232
         8.3.2.1 The Text/Document Editor 232
      8.3.3 Reading data into NLOGIT 232
      8.3.4 Writing data into NLOGIT 235
      8.3.5 Saving data sets 235
9 Case study: mode-choice data

9.1 Introduction
9.2 Study objectives
9.3 The pilot study
  9.3.1 Pilot sample collection
    9.3.1.1 Interviewer briefing
    9.3.1.2 Interviewing
    9.3.1.3 Analysis of contacts
    9.3.1.4 Interviewer debriefing
9.4 The main survey
  9.4.1 The mode-choice experiment
    9.4.1.1 Detailed description of attributes
    9.4.1.2 Using the showcards
  9.4.2 RP data
  9.4.3 The household questionnaire
  9.4.4 The commuter questionnaire
  9.4.5 The sample
    9.4.5.1 Screening respondents
    9.4.5.2 Interviewer briefing
    9.4.5.3 Interviewing
    9.4.5.4 Analysis of total contacts
    9.4.5.5 Questionnaire check edit
    9.4.5.6 Coding and check edit
    9.4.5.7 Data entry
    9.4.5.8 SPSS setup
9.5 The case study data
  9.5.1 Formatting data in NLOGIT
  9.5.2 Getting to know and cleaning the data
Appendix 9A The contextual statement associated with the travel choice experiment
Appendix 9B Mode-choice case study data dictionary
Appendix 9C Mode-choice case study variable labels

10 Getting started modeling: the basic MNL model

10.1 Introduction
10.2 Modeling choice in NLOGIT: the MNL command
10.3 Interpreting the MNL model output
   10.3.1 Maximum likelihood estimation
   10.3.2 Determining the sample size and weighting criteria used
   10.3.3 Interpreting the number of iterations to model convergence
   10.3.4 Determining overall model significance
   10.3.5 Comparing two models
   10.3.6 Determining model fit: the pseudo-$R^2$
   10.3.7 Type of response and bad data
   10.3.8 Obtaining estimates of the indirect utility functions
      10.3.8.1 Matrix: LastData/LastOutput
   10.4 Interpreting parameters for effects and dummy coded variables
   10.5 Handling interactions in choice models
   10.6 Measures of willingness to pay
   10.7 Obtaining choice probabilities for the sample
   10.8 Obtaining the utility estimates for the sample
Appendix 10A Handling unlabelled experiments

11 Getting more from your model
11.1 Introduction
11.2 Adding to our understanding of the data
   11.2.1 Show
   11.2.2 Descriptives
   11.2.3 Crosstab
11.3 Adding to our understanding of the model parameters
   11.3.1 Effects: elasticities
   11.3.2 Calculating arc elasticities
   11.3.3 Effects: marginal effects
11.4 Simulation
   11.4.1 Marginal effects for categorical coded variables
   11.4.2 Reporting marginal effects
11.5 Weighting
   11.5.1 Endogenous weighting
   11.5.2 Weighting on an exogenous variable
11.6 Calibrating the alternative-specific constants of choice models
   estimated on SP data
   11.6.1 Example (1) (the market shares of all alternatives are known a priori)
   11.6.2 Example (2) (the market shares for some alternatives are unknown)
Appendix 11A Calculating arc elasticities

12 Practical issues in the application of choice models
12.1 Introduction
12.2 Calibration of a choice model for base and forecast years
12.3 Designing a population data base: synthetic observations
12.4 The concept of synthetic households
  12.4.1 Synthetic households’ generation framework 441
12.5 The population profiler
  12.5.1 The synthetic household specification 444
12.6 The sample profiler 448
12.7 Establishing attribute levels associated with choice alternatives in the base year and in a forecast year 451
12.8 Bringing the components together in the application phase 452
12.9 Developing a decision support system
  12.9.1 Using the data sample averages in creating the DSS 455
  12.9.2 Using the choice data in creating the DSS 461
  12.9.3 Improving the look of the DSS 472
  12.9.4 Using the DSS 475
12.10 Conclusion 475

Part II Advanced topics

13 Allowing for similarity of alternatives 479
  13.1 Introduction 479
  13.2 Moving away from IID between all alternatives 481
  13.3 Setting out the key relationships for establishing a nested logit model 482
  13.4 The importance of a behaviorally meaningful linkage mechanism between the branches on a nested structure 486
  13.5 The scale parameter 487
  13.6 Bounded range for the IV parameter 490
  13.7 Searching for the “best” tree structure 494
Appendix 13A Technical details of the nested logit model 495

14 Nested logit estimation 518
  14.1 Introduction 518
  14.2 The Hausman-test of the IIA assumption 519
  14.3 The nested logit model commands
    14.3.1 Normalizing and constraining IV parameters 534
    14.3.2 RU1 and RU2 538
    14.3.3 Specifying start values for the NL model 538
    14.3.4 A quick review of the NL model 539
  14.4 Estimating an NL model and interpreting the output
    14.4.1 Estimating the probabilities of a two-level NL model 549
    14.4.2 Comparing RU1 to RU2 556
  14.5 Specifying utility functions at higher levels of the NL tree 564
  14.6 Handling degenerate branches in NL models 570
  14.7 Three-level NL models 574
  14.8 Searching for the best NL tree structure: the degenerate nested logit 577
  14.9 Combining sources of data: SP-RP 580
15 The mixed logit model
15.1 Introduction 605
15.2 Mixed logit choice models 606
15.3 Conditional distribution for sub-populations with common choices 610
15.4 Model specification issues 611
  15.4.1 Selecting the random parameters 611
  15.4.2 Selecting the distribution of the random parameters 612
  15.4.3 Imposing constraints on a distribution 614
  15.4.4 Selecting the number of points for the simulations 614
  15.4.5 Preference heterogeneity around the mean of a random parameter 617
  15.4.6 Accounting for observations drawn from the same individual correlated choice situations 617
  15.4.7 Accounting for correlation between parameters 619
15.5 Willingness-to-pay challenges 620
15.6 Conclusions 621

16 Mixed logit estimation 623
16.1 Introduction 623
16.2 The mixed logit model basic commands 623
16.3 NLOGIT output: interpreting the mixed logit model 627
16.4 How can we use random parameter estimates? 635
  16.4.1 A note on using the lognormal distribution 641
16.5 Imposing constraints on a distribution 645
16.6 Revealing preference heterogeneity around the mean of a random parameter 650
  16.6.1 Using the non-stochastic distribution 656
  16.6.2 Handling insignificant heterogeneity around the mean parameter estimates 661
16.7 Correlated parameters 667
16.8 Common-choice-specific parameter estimates: conditional parameters 679
16.9 Presenting the distributional outputs graphically using a kernel density estimator 684
16.10 Willingness-to-pay issues and the mixed logit model 686

Glossary 695
References 710
Index 714